Negative Interactions in Irreversible Self-Assembly*

David Doty[†] Lila Kari[‡] Benoît Masson[§]

Abstract

This paper explores the use of negative (i.e., repulsive) interactions in the abstract Tile Assembly Model defined by Winfree. Winfree in his Ph.D. thesis postulated negative interactions to be physically plausible, and Reif, Sahu, and Yin studied them in the context of *reversible* attachment operations. We investigate the power of negative interactions with *irreversible* attachments, and we achieve two main results. Our first result is an impossibility theorem: after t steps of assembly, $\Omega(t)$ tiles will be forever bound to an assembly, unable to detach. Thus negative glue strengths do not afford unlimited power to reuse tiles. Our second result is a positive one: we construct a set of tiles that can simulate an s-space-bounded, t-time-bounded Turing machine, while ensuring that no intermediate assembly grows larger than $O(s)$, rather than $O(s \cdot t)$ as required by the standard Turing machine simulation with tiles. In addition to the space-bounded Turing machine simulation, we show another example application of negative glues: reducing the number of tile types required to assemble "thin" $(n \times o(\log n / \log\log n))$ rectangles.

1 Introduction

Tile-based self-assembly is a model of "algorithmic crystal growth" in which square "tiles" represent molecules that bind to each other via highly-specific bonds on their four sides, driven by random mixing in solution but constrained by the local binding rules of the tile bonds. Erik Winfree [13], based on experimental work of Seeman [10], modified Wang's mathematical model of tiling [12] to add a physically plausible mechanism for growth through time. Winfree defined a model of tile-based self-assembly known as the abstract Tile Assembly Model (aTAM). The fundamental components of this model are un-rotatable, but translatable square "tile types" whose sides are labeled with "glues" representing binding sites. Two tiles that are placed next to each other are attracted with strength determined by the glues where they abut, and in the aTAM, a tile *binds* to an assembly if it is attracted on all matching sides with total strength at least a certain threshold value τ.[1] Assembly begins from a "seed" tile and progresses until no more tiles may attach.

*An extended abstract of this paper appeared as [3]. This research was supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant R2824A01 and the Canada Research Chair Award in Biocomputing to Lila Kari, and by the NSF Computing Innovation Fellowship grant to David Doty.

†California Institute of Technology, Department of Computing and Mathematical Sciences, Pasadena, California, USA, `ddoty@caltech.edu`

‡University of Western Ontario, Department of Computer Science, London, Canada, `lila@csd.uwo.ca`

§IRISA (INRIA), Campus de Beaulieu, Rennes, France, `benoit.masson@irisa.fr`

1The threshold τ models the temperature at which insufficiently strong chemical bonds will break, such as those formed by Watson-Crick complementarity in DNA-based implementations of tiles.

We study a variant of this model in which glue strengths are allowed to be negative as well as positive. This leads to the situation in which a stable assembly may become unstable through the addition of a tile that, while binding strongly enough to the assembly to remain attached itself, exerts a repulsive force on a neighboring tile, which is sufficiently strong to detach some portion of the assembly. This is formally modeled by allowing an assembly to break into two parts any time that the two parts have total connection strength less than τ (i.e., if there is a cut of the interaction graph of strength less than τ). Negative glue strengths were discussed as a plausible mechanism in Winfree's thesis [13], and explored theoretically in a more general model of *graph-based self-assembly* by Reif, Sahu and Yin [7]. We compare the results of [7] to the present paper in more detail later in this section.

This paper has two main contributions, an impossibility result and a positive result. The impossibility result is that under the irreversible model, negative glue strengths are not sufficient to achieve perfect reuse of tiles as in [7]. It is tempting to believe that with negative glue strengths, the monotonic growth of the aTAM could be overcome to such a degree that a bounded set of tiles could be reused for arbitrarily long computations,[2] hence implementing the observation that "you can reuse space but you can't reuse time". Alas, you can't reuse space (tiles) too much with irreversible reactions. We show that under the irreversible model of tile assembly, even with negative glue strengths, many tiles will be forever bound to an assembly, unable to detach. In fact, this number is linear in the number of assembly operations, so that after t operations, $\Omega(t)$ tiles will be permanently bound to some assembly.

The positive result is a construction attempting to make do with this limitation. For concreteness, our construction shows how to simulate a single-tape Turing machine. But the idea applies to the iterated computation of any function f that can be "computed with constant height" by a tile assembly system (a formal definition is given in Section 4). The function f_M mapping the configuration of a Turing machine M to its next configuration is an example of one such function. Other examples include the incrementing or decrementing of a counter, or the selection of a uniformly distributed random number from a finite set $\{1, 2, \ldots, n\}$ using flips of a fair coin via von Neumann's *rejection method*, as shown in [4].

Our construction achieves the following property: if the Turing machine M being simulated on input x (with $n = |x|$) has space bound $s(n)$ and time bound $t(n)$, then $O(t(n) \cdot s(n))$ tiles (meaning total count of tiles, which is greater than the number of unique tile types), mixed in solution, will simulate the computation of M on input x, and no intermediate assembly will grow to size larger than $O(s(n))$. The impossibility result can be interpreted to imply that external energy must be supplied to break bonds between tiles in these intermediate assemblies if we wish to reuse them for computation. If we wish to limit the volume of a solution, and therefore the number of molecules it can contain (by the *finite density constraint*, see [11]) to $O(s(n))$, then we cannot allow intermediate assemblies to grow larger than this value. Of course, by the impossibility result, many more than $s(n)$ different such assemblies will form if $t(n) \gg s(n)$ (for instance, when simulating a linear-space, cubic-time computation). With a mechanism to "vacuum" away junk assemblies and supply the external energy needed to break them up (a mechanism not modeled in the aTAM), these tiles could be reused, bringing down the required number of tiles from $O(t(n) \cdot s(n))$ to $O(s(n))$.

The main difference between [7] and the present paper is that [7] employs *reversible* reactions,

[2] This is subject, of course, to computational complexity constraints such as $\mathsf{DTIME}(t(n)) \subseteq \mathsf{DSPACE}(2^{t(n)})$, based on the observation that configurations cannot repeat during the course of a halting computation.

and the present paper employs *irreversible* reactions.[3] Within the aTAM, the main difference between our model and [7] amounts to a difference in the definition of a legal attachment operation. In [7], the authors define a tile attachment to be legal if the tile attaches with strength $\tau - 1$ (in fact, they define it a bit differently but restricting attention to our construction and that of [7], this definition is equivalent). This is a phenomenon not modeled by the aTAM, but it is physically plausible to suppose that it occurs, though with less frequency than strength τ attachments (see the kinetic TAM of [13]). Therefore the tile may detach after attaching since it is held with insufficient strength. But, if it first causes another tile or group of tiles to be bound with total strength less than τ, then those tiles may also fall off, possibly resulting in stabilization of the original attachment. In the present paper, we define attachments to be legal only if they have strength at least τ, whereas detachments may only happen between assemblies attached with strength at most $\tau - 1$. This difference implies that our impossibility result does not apply to [7], which can be considered an advantage of reversible interactions. But this advantage does not come without disadvantages: due to the second law of thermodynamics, their construction must necessarily be implemented as an unbiased random walk with equal rates of forward and reverse reaction, lest the entropy of the system increase with time if one direction is more favorable. Therefore their construction takes expected time n^2 to go forward n steps.

We should also note that although [7] uses a more general model of graph-based self-assembly, this does not imply that their construction of an assembly system simulating a space-bounded Turing machine simulation is a stronger result than our construction. The more general model affords more power to aid in a construction, such as allowing non-planar interactions, in addition to the extra power of reversible interactions. Therefore, we emphasize that our positive construction is not merely a specialization of the construction of [7] to grid graphs. The construction of [7] is inherently non-planar and reversible.

Another use of negative glue strengths for enhancing the power of "temperature 1" systems was shown by Patitz, Schweller, and Summers [6]. Temperature-1 systems are those in which the binding strength threshold is 1, so that it is not possible (in the standard nonnegative strength model) to implement cooperative binding in which two glues must match before a tile can bind. They show that under a more restricted model of negative glues than we consider, in which 1) all glue strengths are -1, 0, or 1, 2) there is only a single type of negative glue, and 3) only equal glues may interact with non-zero strength, it is possible to achieve Turing universality and assembly of $n \times n$ squares from $O(\log n)$ tile types, the latter result being exponentially better than the best known upper bound of $2n - 1$ tile types [8] in the standard temperature-1 model.

This paper is organized as follows. Section 2 gives an description of the aTAM with negative glue strengths and discusses some issues associated with choosing a proper model of negative interactions in self-assembly. Section 3 states and proves our impossibility result. Section 4 provides the construction for our main positive result. Section 5 shows another simple application of negative glues: reducing the number of tile types required to uniquely assemble a "thin" rectangle (an $n \times k$ rectangle with $k < \log n / (\log \log n - \log \log \log n)$). Section 6 concludes the paper and discusses the utility of negative glue strengths in general.

For color figures, see `http://arxiv.org/abs/1002.2746`.

[3] [7] also uses a more general graph-based model of self-assembly, but this difference is less crucial than the reversibility issue.

2 Abstract Tile Assembly Model

This section gives a brief definition of the abstract Tile Assembly Model (aTAM, [13]) with negative glue strengths. This not a tutorial on the aTAM; for readers unfamiliar with the model, please see [8] for an excellent introduction.

2.1 Issues with Choice of Model

There are many variations on the model of tile self-assembly with negative glue strengths. We identify six (somewhat) independent binary choices to be made.

Seeded/Unseeded. In the seeded model, assembly starts from a specially designated seed tile. In the unseeded model, assembly may start with any tile.

Single-tile addition/Hierarchical assembly. In the single-tile addition model, tiles are added one at a time to an assembly. In hierarchical assembly [1], two assemblies, each possibly consisting of multiple tiles, may attach to each other.

Irreversible/Reversible. In the irreversible model, for a tile/assembly to attach to an assembly, it must bind with strength $\geq \tau$ (though it may cause another cut of the resulting assembly to have strength $< \tau$). In the reversible model (used by Reif, Sahu, and Yin [7]), a tile/assembly may attach with strength $< \tau$, implying that it may detach (reverse), but may also cause another cut to detach. In [7], most attachment events have the property that they cause precisely two cuts to have strength $\tau - 1$, that of the attachment event itself, and another desired cut, and if the latter cut detaches, then the former cut now has strength τ. Hence, on the assumption that each cut is equally likely, assembly proceeds in a unbiased random walk (thus taking expected time n^2 to go forward n steps).

Detachment precedes attachment/Detachment and attachment in arbitrary order. If detachment precedes attachment, this means we assume that whenever a negative-strength glue creates an unstable cut, every detachment event that can occur, will occur, until we are left with nothing but stable assemblies, before the next attachment event occurs. The more realistic model, detachment and attachment in arbitrary order, assumes that the operations of attach and detach are both legal at any stage. Therefore, if a negative-strength glue creates an unstable cut, *but* this cut could be stabilized if an attachment secures it in place before it can detach, then we assume that this could happen.

Finite tile counts/Infinite tile counts. Most models of tile self-assembly assume an infinite number of tile types, since a very large number can be easily created in a short time. However, if we wish to use negative glue strengths to implement space-bounded computation, and the finite density constraint (see [11]) implies that the solution volume must be proportional to the number of molecules it contains, then we must constrain the number of molecules. See the next point for a discussion of another practical issue with using bounded space for computation with tile assembly.

Tagged result/Tagged junk. This choice relates to how we designate what is the "result" of assembly. In the tagged result model, we designate a subset of tile types to be "black", and state that a result assembly is any terminal assembly with a black tile in it. This allows us to separate the junk from the result after assembly is complete, but does not allow junk to be removed *during* assembly, since the black tile may not be attached to the result assembly until the very end of the assembly process. In the tagged junk model, we say that any assembly (terminal or not) with a black tile is junk and may be removed.[4] Furthermore, every producible assembly must have the property that it may either grow into the result, or will grow into an assembly tagged with a black tile as junk. That is, at any point in the assembly process, we could "vacuum" away all assemblies tagged with a black tile, knowing that anything remaining is not (yet) junk and can be re-used. This allows a computation with space bound s and running time $t \gg s$ to be done in volume $O(s)$, so long as one uses detachment to ensure that no intermediate assembly grows larger than $O(s)$, by supplying just enough copies of tiles that are not swept away, and periodically sweeping away the tagged junk while supplying fresh unattached tiles to carry out more work.[5]

Note that neither of these tagging models are typically used in other tile self-assembly papers, but the special implications of negative glue strengths imply that we cannot simply follow the convention that the seed identifies the result, as discussed below.

Of these, the first three choices are incomparable in terms of power: each choice affords both advantages and disadvantages in terms of designing a tile assembly system. The latter three choices are more clear: for each choice, one option makes implementing a correct design strictly more difficult, but results in a more robust and realistically implementable construction. Respectively, these choices are 1) detachment and attachment in arbitrary order, 2) finite tile counts, and 3) tagged junk.

These choices are not completely independent. For instance, there are two (mathematical) uses for a seed tile: 1) to identify *legal attachment events in single-tile addition*: attachment is legal if it is between a single tile and an assembly containing the seed, and 2) to identify the *result of assembly with hierarchical growth*: one must allow attachment of tiles separate from the seed, but identifies legal results as terminal assemblies containing the seed. We separate out these uses by making the method of "tagging" the result independent of the seed tile, using the seed only to define which assemblies are producible. With negative glues implying the ability to remove the seed from an assembly, we must take care to allow attachment events in the seeded model not only with assemblies containing the seed, but with those *derived from the seed*.

For this paper we use the model of single-tile addition, irreversible, seeded, detachment and attachment in arbitrary order, infinite tile counts, and tagged result.

[4] Actually it is more realistic to assume something like *two* different black tiles, placed at a certain recognizable orientation (such as next to each other) to enforce that individual black tiles in solution are not interpreted as identifying junk until they actually attach to something, but we do not dwell on this issue since we use the tagged result model.

[5] To implement such an idea, it may be necessary to obey nontrivial constraints on the tile type composition of the junk assemblies. For instance, if a Turing machine simulation with tape alphabet $\{a, b, c, d\}$ has all a and b characters on the tape for a long time period $[t_1, t_2]$, then all c and d characters on the tape for another long time period $[t_3, t_4]$, then the experimenter would need to supply a and b tile types in time period $[t_1, t_2]$ but not c and d tile types until time period $[t_3, t_4]$, to avoid filling the tube with excess of the tile types that are not being used.

2.2 Definition of Model

$\mathbb{Z}$ and $\mathbb{Z}^+$ denote the set of integers and positive integers, respectively. Let G be a finite alphabet of *glues*. A *tile type* is a tuple $t \in G^4$, i.e., a unit square with a glue on each side. Associated with the tile types is a *glue strength function* $str : G \times G \to \mathbb{Z}$ that indicates, given two glues g_1 and g_2, the strength $str(g_1, g_2)$ with which they interact. Unlike the standard model [8, 13], our glue strength function is *non-diagonal*, meaning that we allow $str(g_1, g_2) \neq 0$ even when $g_1 \neq g_2$. We assume a finite set T of tile types, but an infinite number of copies of each tile type, each copy referred to as a *tile*. Let $G(T)$ denote the set of all glues of tile types in T. An *assembly* (a.k.a., *supertile*) is a positioning of tiles on the integer lattice $\mathbb{Z}^2$ (i.e., a partial function $\alpha : \mathbb{Z}^2 \dashrightarrow T$, where $\dashrightarrow$ denotes that the function is partial). Each assembly induces a *binding graph*, a grid graph whose vertices are tiles, with an edge between two tiles if they are adjacent (i.e., are Euclidean distance 1 apart).[6] The assembly is τ*-stable*, or simply *stable* if τ is understood from context, if every cut of its binding graph has weight (strength) at least τ, where the weight of an edge is the strength of the glue it represents. That is, the assembly is stable if at least energy τ is required to separate the assembly into two parts. In this paper, where not stated otherwise, we assume that $\tau = 2$.

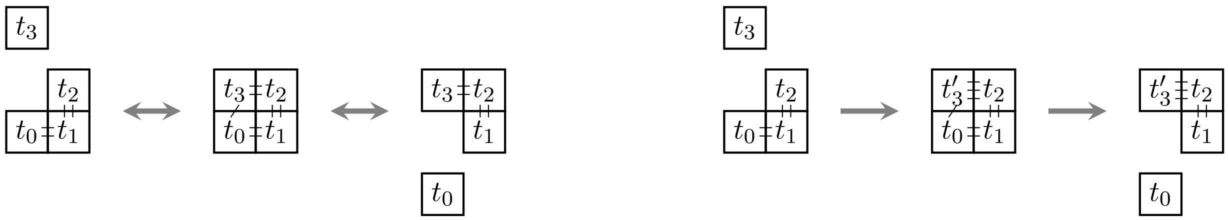
A *tile assembly system* (TAS) is a 4-tuple $\mathcal{T} = (T, str, \sigma, \tau)$, where T is a finite set of tile types, $str : G(T) \times G(T) \to \mathbb{Z}$ is the *glue strength function*, $\sigma : \mathbb{Z}^2 \dashrightarrow T$ is the finite and τ-stable *seed assembly*, and $\tau \in \mathbb{Z}^+$ is the *temperature*. Given a TAS $\mathcal{T} = (T, str, \sigma, \tau)$, an assembly α is *producible* if either (base case) $\alpha = \sigma$, or (recursive case 1) α results from the τ-stable attachment of a single tile to a producible assembly ("τ-stable attachment" meaning that the cut separating the tile from the rest of the assembly has strength $\geq \tau$), or (recursive case 2) α consists of one side of a cut of strength $< \tau$ of a producible assembly. Note in particular that a producible assembly need not be stable, but may be stabilized by attachments before it can break apart. An assembly α is *terminal* if α is τ-stable (so no detachments are possible) and no tile can be τ-stably attached to α. To formally define our notion of "tagged result" from Section 2.1, let $B \subseteq T$ be a set of "painted black" tile types. $\mathcal{T}$ is *B-directed* (a.k.a., *B-deterministic*, *B-confluent*) if it has exactly one terminal, producible assembly containing one or more tiles from B.[7]

To define reversible assembly at temperature $\tau = 2$ (as in [7]), it suffices to define attachment events with strength threshold $\tau - 1 = 1$, rather than strength threshold $\tau = 2$. This behavior is illustrated in Figure 1(a), and can be compared with our new notion, whose evolution is shown in Figure 1(b).

To define unseeded assembly, it suffices to drop σ from the definition of TAS, and define the base case of a producible assembly as any individual tile. To define *hierarchical* assembly (a.k.a., *two-handed*, *multiple tile* [1]), it suffices to change the first recursive case to state that legal attachment events are between any two producible assemblies, such that they can be positioned in such a way that the cut separating them has strength $\geq \tau$ (i.e., can be stably attached). Then, an assembly α is *terminal* if it is τ-stable and for every producible assembly β, α and β cannot be stably attached. Figure 2 illustrates the new behaviors allowed by the hierarchical variant.

[6] Previous papers model the binding graph as having edges only between tiles that interact with positive strength. In the present paper, the presence of negative glue strengths means that we must consider every possible interaction between adjacent tiles, whether positive, negative, or 0.

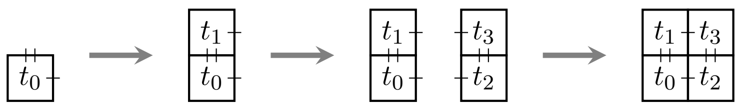
[7] We define this notion of B-directedness but do not henceforth discuss it explicitly, since our main construction simulates a general "computation", and B would depend on the goals of the computation being simulated. In our example construction in Section 4 of simulating steps of a Turing machine, B could, for instance, consist of the tile types that represent a halting state, so that only a terminal assembly representing the configuration of a halted Turing machine would be considered the result.



(a) Reversible model, as defined in [7].

(b) Irreversible model.

Figure 1: Two different implementations of negative interactions at temperature 2. The slanted bonds represent a strength of -1. In the reversible model, the tile t_3 can attach with a total strength of 1 (one bond of strength 2 and one of strength -1) and hence is unstable, while with our definition, t_3' is attached with a total strength of 2 and forces t_0 to detach.



(a) (b) (c) (d)

Figure 2: Typical example of hierarchical assembly, at temperature $\tau = 2$. The segments between tiles represent the bonds, the number of segments encodes the strength of the bond (here, 1 or 2). In the seeded, single tile model with seed $\sigma = t_0$, the assembly at step (b) would be terminal.

3 Limitation of Tile Reuse with Irreversible Reactions

If α is an assembly and $p, q \in \mathrm{dom}\ \alpha$ are two adjacent positions in α, with glues g_p and g_q touching where the tiles $\alpha(p)$ and $\alpha(q)$ meet, define $g_\alpha(p,q) = str(g_p, g_q)$ to be the strength of their interaction. Let $g_\alpha(p,q) = 0$ if p and q are not adjacent positions. Define $\Phi(\alpha) = \frac{1}{2}\sum_{p,q\in \mathrm{dom}\ \alpha} g_\alpha(p,q)$, the *(negative) free energy* of α, to be the sum of all glue strengths between adjacent tiles in the assembly (the fraction $\frac{1}{2}$ corrects for the double counting of pairs of positions).[8] In particular, an assembly consisting of a single tile has free energy 0. If S is a multiset of assemblies (such as that produced by a TAS with negative glue strengths, considering even the "junk" assemblies that are discarded after a cut), define the (negative) free energy of S to be the sum of the free energies of each assembly in S, denoted $\Phi(S)$. Note that even postulating an infinite count of tiles, after a finite number of operations, only finitely many assemblies in S consist of more than one tile, and each of these is a finite assembly. Therefore $\Phi(S) < \infty$ for any multiset S of assemblies producible by a TAS, even in the case that $|S| = \infty$ (such as the initial multiset consisting of a countably infinite number of copies of each individual tile type).

When we discuss the "number of steps" for the assembly process of a TAS, we mean the total number of attachment and detachment operations. We do not claim that this is a proper model of "running time", but it is convenient to think of attachment and detachment events as discrete and equally-spaced steps, even though they may happen in parallel or with interval times governed by

[8]The standard definition of free energy is the negative of this quantity, but as in [8] we use its negation so that the quantity will be positive for stable assemblies. Intuitively, it is the energy *required* to separate α into individual tiles, whereas the standard definition is the energy *released* by such a separation.

a continuous distribution.

Theorem 3.1. *Let $\mathcal{T}$ be a TAS, and let S be a multiset of assemblies producible by $\mathcal{T}$ after $t \in \mathbb{N}$ steps. Then $\Phi(S) \geq t/2$.*

Proof. Suppose that $\mathcal{T}$ has a seed of size 1; otherwise, the free energy we derive for step t will be even higher, so this assumption does not harm the proof. For $i \in \{0, 1, \ldots, t\}$, let S_i denote the multiset of assemblies after the first i operations, so that $S_t = S$ and S_0 is the multiset of individual unattached tiles (with a countably infinite number of copies of each tile type). Note that $\Phi(S_0) = 0$. Let $A \subseteq \{1, 2, \ldots, t\}$ be the indices of attachment operations in the first t operations, and let $D = \{1, 2, \ldots, t\} - A$ be the indices of detachment operations, so that operation i changes S_{i-1} to S_i.

Each attachment operation increases the free energy by at least τ for a system operating at temperature τ, since we require a tile attachment to have the property that the cut between the tile and the rest of the assembly has strength at least τ, and the edges of this cut previously each contributed 0 to the free energy since they were all unbound. So for S_{i-1} leading to S_i via attachment, $\Phi(S_i) - \Phi(S_{i-1}) \geq \tau$. For each detachment operation, the greatest strength cut that could be broken to create the detachment has strength $\tau - 1$; stronger cuts cannot be broken. This implies the free energy decreases by at most $\tau - 1$ during a single detachment operation.[9] So for S_{i-1} leading to S_i via detachment, $\Phi(S_i) - \Phi(S_{i-1}) \geq -(\tau - 1)$. Amortizing over all operations,[10] we have that

$$\begin{aligned}
\Phi(S) &= \sum_{i=1}^{t} (\Phi(S_i) - \Phi(S_{i-1})) \\
&= \sum_{i \in A} (\Phi(S_i) - \Phi(S_{i-1})) + \sum_{i \in D} (\Phi(S_i) - \Phi(S_{i-1})) \\
&\geq \sum_{i \in A} \tau + \sum_{i \in D} -(\tau - 1) \\
&= |A|\tau - |D|(\tau - 1).
\end{aligned}$$

Although we posit an infinite number of copies of each tile type, during the first t steps at most $t+1$ tiles, denoted as the multiset S'_0, can actually participate in assembly operations. Let c_i denote the total number of assemblies in S_i that consist of tiles from S'_0, so that c_0 is simply the total number of initial tiles that will participate in the first t steps. Each attachment event decreases c_i by one, and each detachment event increases c_i by one. Since $c_i \leq c_0$ for all i (we cannot have more assemblies than there are tiles), this implies that for all i, the number of attachment events in the first i steps is at least the number of detachment events in the first i steps. Therefore $|A| \geq |D|$ and since $|A| + |D| = t$, we conclude $|A| \geq t/2$, whence the above inequality tells us that

[9] A single attachment of a tile with negative glue strength can potentially cause a cascade of detachments that, put together, lead to a large decrease in free energy. However, these are each considered separate detachment events.

[10] See [2, Section 17.3] for a discussion of amortized analysis, which is a fancy phrase for writing the following sum in this form.

$$
\begin{aligned}
\Phi(S) &\geq |A|\tau - |D|(\tau - 1) \\
&\geq |A|\tau - |A|(\tau - 1) \\
&= |A| \\
&\geq t/2. \qquad \square
\end{aligned}
$$

The proof works for hierarchical systems, and since single-tile addition systems are simply hierarchical systems with an extra constraint on legal attachment operations, the proof applies to single-tile addition systems as well. The proof also works both for seeded and unseeded systems. As there is no property of the 2D plane or grid graphs employed in our proof, the proof applies to the *irreversible* version of the graph-based self-assembly model studied in a reversible context by Reif, Sahu, and Yin [7].

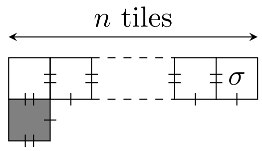
Since the glue strengths in any given tile system have some maximum value $s \in \mathbb{Z}^+$, an immediate consequence of Theorem 3.1 is that after t steps, at least $t/(2s)$ sides of tiles are bound. With the finite tile count assumption, once t is sufficiently large that $t/(2s)$ exceeds the total number of sides available (i.e., 4 times the total number of tiles in solution), no more sides are available for binding, and self-assembly necessarily grinds to a halt. This is the sense in which a finite number of tiles cannot be reused indefinitely.

An interesting question is how tight Theorem 3.1 can be in general. For example, is there a tile system $\mathcal{T}$ that for infinitely many $t \in \mathbb{N}$ reaches a multiset of assemblies S_t such that $\Phi(S_t) \leq t/2 + o(t)$?

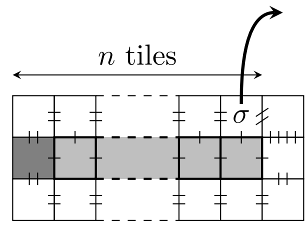
Some seemingly straightforward attempts to prove Theorem 3.1 fail in ways that illustrate potentially nonintuitive properties of negative glue strengths. It is not true, for instance, that the free energy increases monotonically, since it drops whenever a cut of positive strength is detached, so a straightforward inductive argument fails. Furthermore, it is not even true that the free energy decreases by at most a constant in between consecutive periods of increase. Even with fixed glue strengths (4 suffices), for each $n \in \mathbb{N}$, it is possible to construct a tile set with the property that there are two states of the system S_i and S_j, with state S_i preceding S_j, such that $\Phi(S_i) \geq \Phi(S_j) + n$. But, Theorem 3.1 implies that any attempt to create such a cascade of detachments that drops the free energy by n requires first attaching even stronger – and ultimately unbreakable – bonds required to set up the state S_i. That is, the free energy can fall arbitrarily far, but in order to do so it must first climb more than twice as high as it will eventually fall. This phenomenon is illustrated in Figure 3, where the last stable addition of a tile leads to an arbitrary decrease of the free energy. This was made possible by the use of stronger strength-4 bonds prior to this event.

There is a natural thermodynamic interpretation of Theorem 3.1: work done by tiles on tiles, in an irreversible manner, increases the entropy of the system by the second law of thermodynamics, thus decreasing the potential energy available to do more work. Therefore, any potential energy stored in the unattached glues is eventually permanently used up if external energy is not supplied to break these bonds. In our main construction, many junk assemblies are created that are no longer useful once the tiles in them have been used once. Theorem 3.1 tells us that no amount of cleverness will allow us to break up those assemblies and reuse the tiles solely through design of tile types with negative glues; some external force must be supplied to break them apart using a mechanism not modeled in the aTAM.

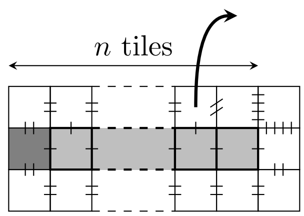
Of course, Theorem 3.1, interpreted in light of the molecular interactions that are being modeled by the aTAM, should not be surprising to any physicist. But we believe it is important to formally



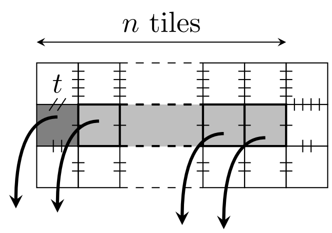
(a) Starting from the seed σ, $n-1$ tiles attach on the left, before a gray tile can be attached at the bottom.



(b) From the gray tile, $n-1$ light gray tiles attach on the right, using the bonds on the top and on the left. The assembly also goes around these new tiles counter-clockwise, until σ is reached and forced to detach.



(c) Starting from σ, the initial tiles are removed by negative bonds of strength -2, and replaced by n new tiles.



(d) Once tile t is attached, the final step is the detachment of the gray tile, followed by the detachment of the $n-1$ light gray tiles one after the other.

Figure 3: A possible evolution where the stable addition of one tile (marked t) can lead to n tiles detaching one after the other (black arrows), hence reducing the free energy by n. As usual, the strength of bonds is represented by the number of segments between tiles, slanted bonds indicating a negative strength. Note that this is only one among many possible evolutions, since there may be several cuts of strength lower than τ that can be removed. In particular, at the last step, all the gray and light gray tiles can detach as one unique big cut, which will in turn break into pieces.

establish the truth of such a statement within the model. One develops more confidence in a model of reality when it tells us something already known about reality (e.g., the Positive Mass Theorem [9]).

Theorem 3.1 does not apply to the negative glue strength construction of Reif, Sahu, and Yin [7], because their model allows *reversible* reactions. Attempting to apply our proof to their model would result in the first inequality $\Phi(S_{i+1}) - \Phi(S_i) \geq \tau$ being replaced by $\Phi(S_{i+1}) - \Phi(S_i) \geq \tau - 1$, which would result in a final lower bound of 0, instead of $t/2$, for $\Phi(S)$. Intuitively, the reversibility of reactions implies that attachment and detachment have symmetric effects on the free energy. But this also implies that their system requires driving the system forward through an unbiased random walk, taking n^2 steps on average to proceed by n net forward steps. Any attempt to speed up the reaction to make the forward rate of reaction faster than the reverse rate of reaction would introduce the imbalance in their respective effects on free energy that allows our proof to work. Therefore this tradeoff in speed versus reusability of tiles is fundamental.

4 Turing Machine Simulation

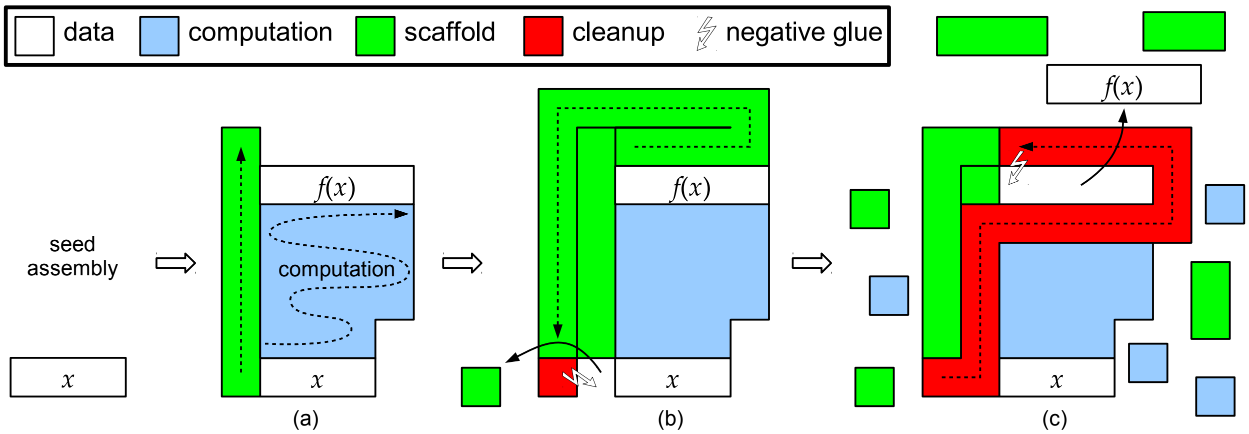


Figure 4: High-level overview of assembly for computation of constant-row computable function f.

Throughout this section, fix some finite alphabet Σ. This section describes the main construction of this paper, that of a simulation of a s-space-bounded Turing machine with a tile system that allows no assembly to grow larger than size $O(s)$. The actual construction is a bit more general, describing the computation of a class of functions, of which the transition function of a Turing machine is one example. Intuitively, the class of functions are those computable by a constant number of rows of assembly (although the number of columns is unbounded) in the standard aTAM. See [8], for a formal definition of the standard aTAM model with nonnegative glue strengths. Briefly, the standard aTAM is the same as the model defined in Section 2, but glue strengths are non-negative and are only positive between equal glues.

Definition 4.1. Let T be a set of tile types, and let $e : T \to \Sigma$. We say that a row of tiles (a connected subassembly of some assembly of height 1) $t_1, t_2, \ldots, t_k$ *e-encodes* a string $x \in \Sigma^k$ if $e(t_1) = x[1], e(t_2) = x[2], \ldots, e(t_k) = x[k]$, where $x[i] \in \Sigma$ is the i^{th} symbol in x. A function $f : \Sigma^* \to \Sigma^*$ is *constant-row computable* if there exist a tile set T, a function $e : T \to \Sigma$, and a constant c such that, for each $x \in \Sigma^*$, there is a height-1 stable assembly $\sigma_x : \mathbb{Z}^2 \dashrightarrow T$ e-encoding x such that the tile assembly system $\mathcal{T} = (T, str, \sigma_x, 2)$ (with $str(g_1, g_2) > 0 \iff g_1 = g_2$) has the unique terminal assembly α, the height of α is c, the bottom row of α is σ_x, the top row of α e-encodes $f(x)$, and the leftmost column of any row of α is no further left than the bottom row.

The widths of the rows representing the input and output may be different (i.e., possibly $|x| \neq |f(x)|$). In this case, we require only that the leftmost and rightmost tiles of each row have their glues specially marked to distinguish them from the tile types interior to the row.

Our construction shows how to design a tile set that will compute iterations of any constant-row computable function f, ensuring that no intermediate assembly grows larger than the size of the input or output processed by any *individual* invocation of f. Examples of such functions include the function f that, given a configuration of a single-tape Turing machine outputs the next configuration of this Turing machine, or that increments a counter represented in binary.

Figure 4 shows a high-level overview of the entire construction, in terms of a general constant-row computable function f. For concreteness, think of f as the function that, given a configuration

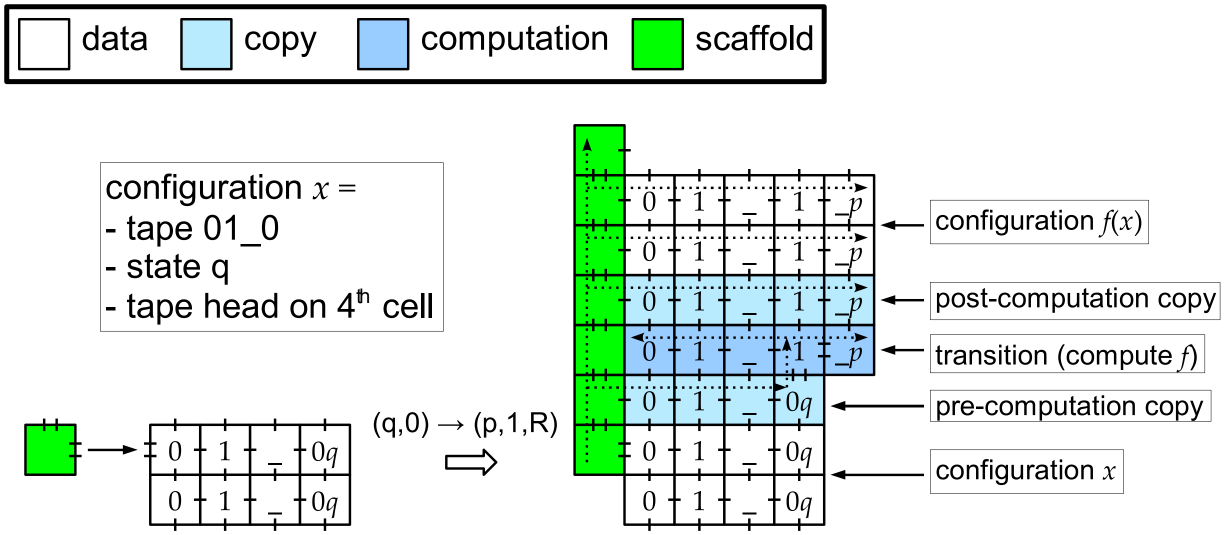


Figure 5: Example of tiles implementing the computation step. Arrows within tiles show order of growth. In this case f is constant-row computable with constant $c = 1$. The first and last copy rows, shown in lighter shade than the center computation tiles, are always present no matter the function f, and their placement is initiated by the scaffold tiles. However, there is no interaction between the center computation and scaffold tiles. Note that the data tiles are two rows with strength 1 glues; this is to make them stable at temperature 2 but not producible (without additional scaffolding) as they would be if they were a single row connected with strength 2 glues.

of a t-time-bounded, s-space-bounded, single-tape Turing machine, outputs the next configuration of this Turing machine (extending the tape on the right side only). The construction proceeds as follows, each label corresponds to a picture in Figure 4.

(a) First, the scaffold tiles (green) connect to the x data assembly (white). The scaffold tiles initiate the computation of f (blue).

(b) The scaffold "detects" when the computation is finished, in the sense that the green row above $f(x)$ tiles cannot complete until all of $f(x)$ is present. Then the scaffold tiles grow back to the first scaffold tile to initiate the removal of $f(x)$ from the tiles surrounding $f(x)$.

(c) Each cleanup tile (red) uses a negative glue strength against the tile "in front of" (on the path show by the arrows) the cleanup tile, and once this tile is removed, a new cleanup tile grows in its place to continue the cleanup. The path and bond placements and strengths are carefully chosen to ensure that no portion of $f(x)$ is removed, until the last step when $f(x)$ detaches whole from the rest of the tiles.

Note that since f is constant-row computable, the height of the scaffold and cleanup parts are bounded by a constant and therefore may be hard-coded into the tile set, whereas special glues mark the horizontal endpoints so that a constant set of tiles may be used for the whole horizontal length of x and $f(x)$, without constraining their length to be constant.

The simulation of the Turing machine for t steps will then consist of executing this assembly process for t iterations, using the output assembly $f(x)$ of each iteration as the input assembly x for

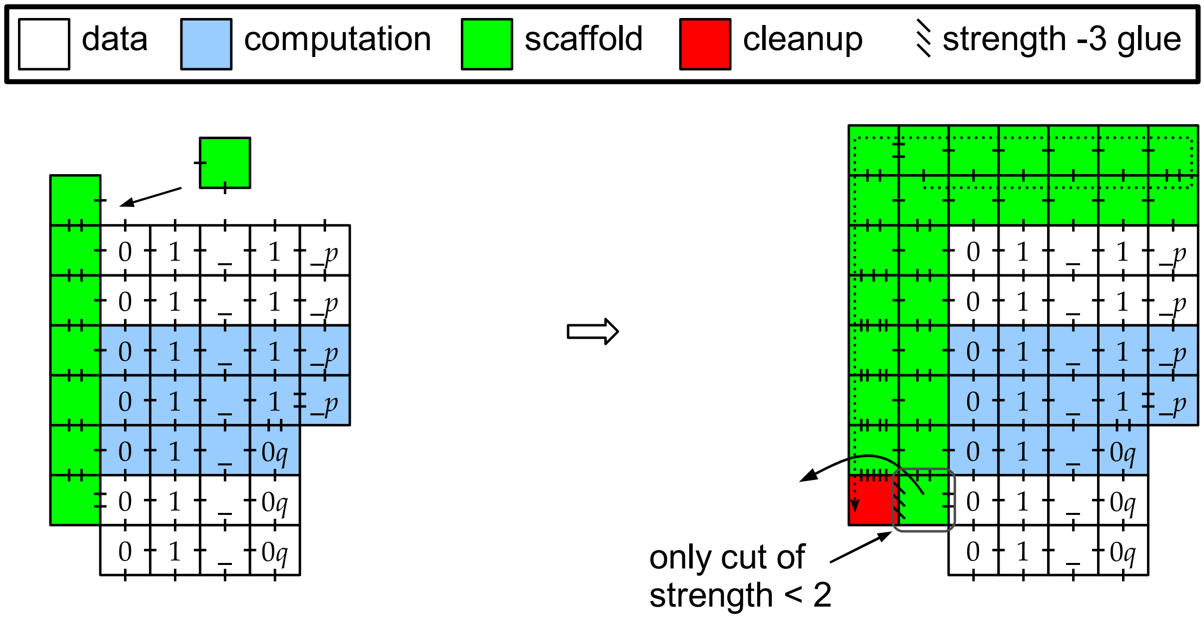


Figure 6: Tiles that position the cleanup tiles. Here the "copy" tiles from Figure 5 are depicted in the same shade as the computation tiles; now that $f(x)$ has been computed our goal is to remove all of them from the subassembly representing $f(x)$. The order of growth of the scaffold tiles ensures that cleanup does not begin until all of $f(x)$ is present.

the next iteration. After each iteration, the width of the remaining "junk" assembly is a constant plus $O(1) + \max\{|x|, |f(x)|\}$, and the height is constant since f is constant-row computable, so the size of the intermediate assemblies is $O(\max\{|x|, |f(x)|\})$.

Figures 5, 6, and 7 give some more details for the three main steps of Figure 4, respectively (a), (b), and (c), using the specific example of f mapping a configuration of a single-tape Turing machine to its next configuration.

Figure 5 shows an example of tiles implementing step (a) of Figure 4, i.e., the computation of f. The example shows one transition of a single-tape Turing machine, with tape contents 01_0 (_ standing for blank), in state q, with tape head on the rightmost cell, transitioning to state p, moving the tape head right, changing the cell's symbol from 0 to 1, and encountering a blank on the new rightmost cell. In this case, a new rightmost cell is needed, illustrating how our construction handles dynamically changing space requirements, but if the tape head were further left in the row, it would simply fill in copy tiles to the right, just as to the left as shown above, and the row would stay the same width. At the start and end of a computation, the configuration is copied so that any strength > 1 bonds used in the computation are on the interior of the computation tiles, ensuring that only strength-1 bonds must later be broken to separate the data tiles. Each data assembly on either end of the computation tiles is represented by a two-row assembly with only single-strength bonds on its interior, which ensures that when detached, the data assembly will be stable, but that it cannot form on its own without help from the scaffold tiles (which would happen

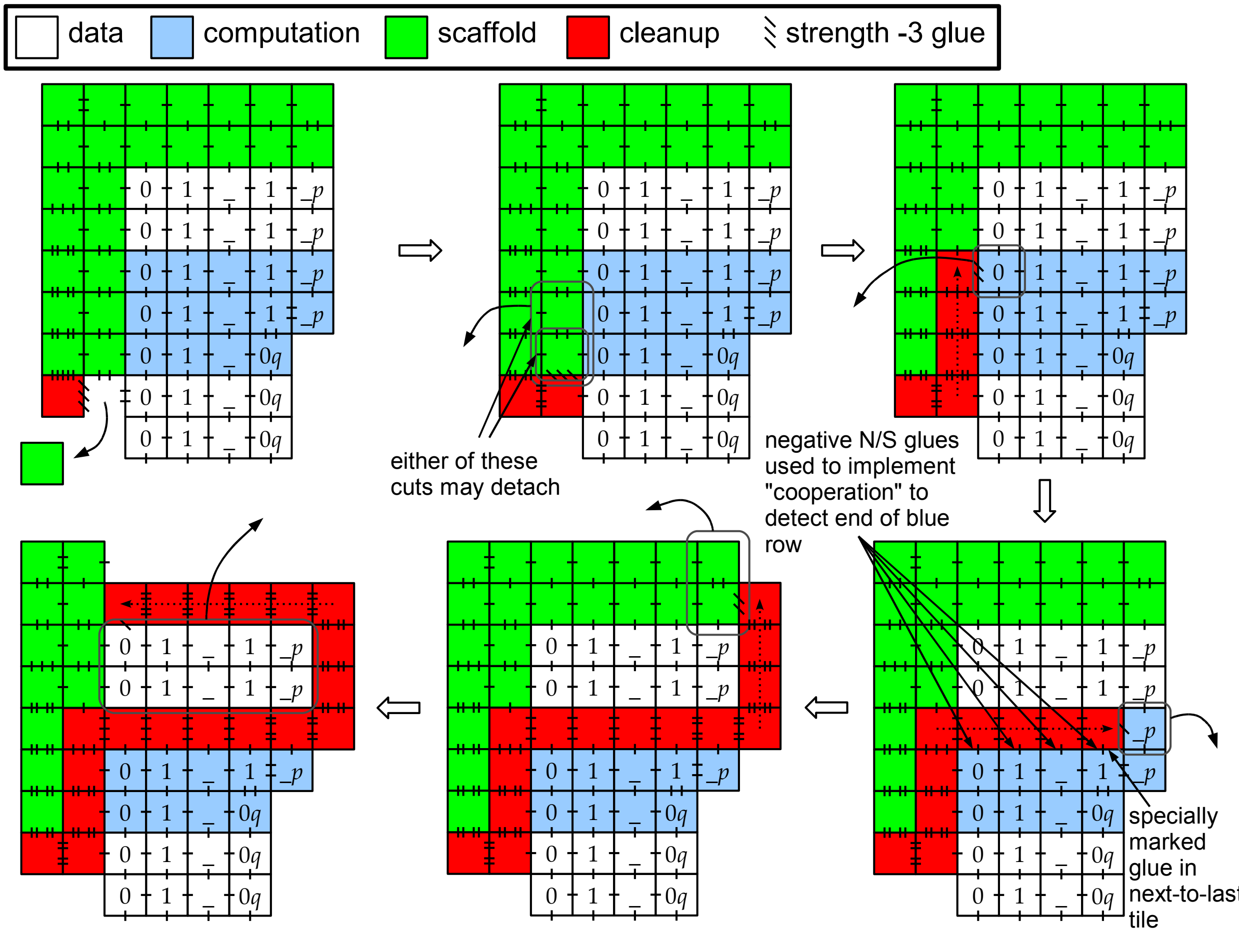


Figure 7: Tiles that "clean up" the connections between the output data and the scaffold and computation tiles to separate them and allow the data tiles to be computed on again. Note that we use the non-diagonality of the strength function at several points. For example, the west glue of the tile to which the scaffold tile of Figure 5 initially binds has strength 2 when binding to the scaffold tile but only strength 1 when binding to the cleanup tiles, to ensure that only the scaffold can bind initially.

if it were only a single row connected with strength-2 bonds). Each vertical position is hard-coded into the tile set; i.e., the scaffold tile set "knows" the required height to compute f. However, the absolute horizontal positions are not encoded into the tiles, only the leftmost and rightmost tiles of the configuration are specially marked, and all interior tile types representing the same data are identical.

Figure 6 shows the tiles implementing step (b) of Figure 4, positioning the tiles for cleanup. The top two rows must use cooperation to tell where is the end of the row underneath, since the width of the output row is unknown. This is necessary to ensure that cleanup does not proceed until the computation of $f(x)$ is complete. The strengths of bonds on the leftmost downwardgrowing column must be sufficiently large to ensure that only the proper cut is made when the first negative-strength glue is applied.

Figure 7 shows the tiles implementing step (c) of Figure 4, "cleaning up" by removing the output $f(x)$ from the scaffold, computation, and x data tiles. Though not shown, negative strength interactions are necessary between the second-to-top row of computation tiles and some of the right-growing cleanup tiles, to ensure that the right end of the row is properly detected. That is, there are two types of cleanup tiles growing right, one to detach the interior tiles, and one to detach the final rightmost computation tile. Since the east-west bonds between cleanup tiles are greater than 1, the negative north-south glue strengths between interior cleanup tiles and the second-to-rightmost blue tile – and between the rightmost cleanup tile and the interior computation tiles – must be strength -2 to ensure that the second-to-rightmost blue tile cannot stably attach except where intended.

By designating the halting tile type(s) as "black", we enforce that the only "result" assembly is the one representing the final configuration of the Turing machine.

5 Reducing Tile Complexity for Thin Rectangles

In this section we show another simple application of negative glue strengths using techniques similar to those used in the construction of Section 4.

Aggarwal, Cheng, Goldwasser, Kao, Espanés, and Schweller [1] studied the tile complexity (number of tile types required to uniquely assembly a shape) of "thin" rectangles. They showed that for all n and all $k < \log n/(\log\log n - \log\log\log n)$, any tile system that uniquely assembles an $n \times k$ rectangle must have at least $\Omega(\frac{n^{1/k}}{k})$ unique tile types. With the model of negative glue strengths (using both the negativity and the non-diagonality of the strength function) we achieve tile complexity $O\left(\sqrt{\log n}\right)$. Since we use a non-diagonal glue strength function, by Theorem 6.2 of [1] (which generalizes from $n \times n$ squares to any shape that encodes the number n), this upper bound is asymptotically optimal.[11] Curiously, it is easier to create a *thick* rectangle than a thin rectangle; we first build a thick rectangle and use negative glues to "cut out" a thinner rectangle of the same length. By placing a "black" tile in the thin rectangle, the tile system is B-directed by the definition given in Section 2, so in this sense the tile system uniquely produces the thin rectangle.

Figure 8 shows the details of this construction. The decoding tile types on the left are identical to the tile types in Figure 5.2 of [1] that encode a natural number n (actually $n/2$; see below) using $O(\sqrt{\log n})$ tile types (rotated clockwise by 90 degrees). The next group of tiles ("copy") create a copy of n to arrange the position of strength-1 glues that will later help to cleanly detach the thin rectangle. The next group of tiles ("n counter") count horizontally eastward from $n/2$ down to 0 (decrementing once every two columns to create n total columns, and using an additional column if the least significant bit of n is 1) to measure the length of the rectangle. Let $m = \lfloor \log(\lfloor \log n \rfloor + 1) \rfloor + 1$. The next group of tiles ("k counter") counts vertically southward from $2^m - k/2$ up to 2^m, to determine the height k at which to start cutting the rectangle. For these k counter tiles the most significant bit is on the right. This can be encoded using $m = O(\log\log n) = o(\sqrt{\log n})$ tile types. Finally, a constant number of tile types, similar to those of Figure 7, cut out the thin rectangle on the bottom. The copy rows between the decoding tiles and the first counter are intended to place strength-1 and strength-2 glues in such a way as to ensure that the final cut separates the thin rectangle, with no additional tiles attached to it. As with the construction of Section 4, we use

[11] We note that both the negativity of the glue strengths and the fact that the glue strength function is non-diagonal are required to achieve this bound.

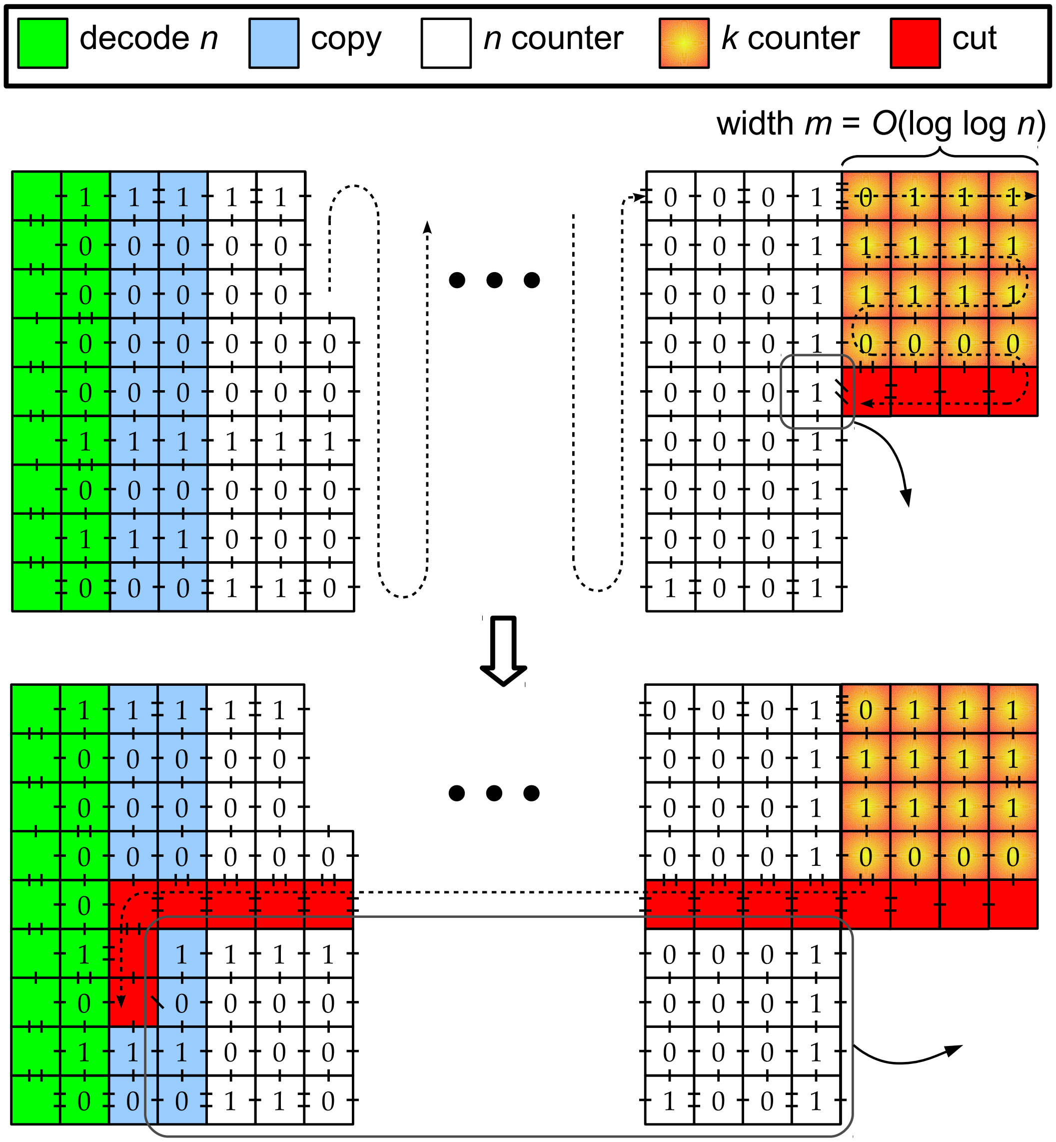


Figure 8: Tile system that produces a thin $n \times k$ rectangle from the asymptotically optimal $O(\sqrt{\log n})$ unique tile types. The different groups of tile types are listed left-to-right in the legend in the order they appear during assembly, and the arrows in the assembly help to illustrated when and where these tiles are placed.

cooperation at the top left of the thin rectangle to allow the red "cut" tiles to know where to turn southward. Despite the fact that the last cut tile in the row above the rectangle attaches to its east neighbor with strength 2, because the north glue has only strength 1 (unlike the previous cut tiles that attached to the north with strength 2), additional cooperative binding strength is required from the west neighbor to overcome the negative glue on the bottom, which is how we switch to south-growing cut tiles.

6 Conclusion

We have shown two main results in the aTAM with negative glue strengths, under the standard assumption of *irreversible* attachment, meaning attachments that only occur with strength at least the temperature τ, versus detachments that can occur on any cut of strength at most $\tau - 1$. The first result is that the amount of tile reuse afforded by the ability to detach tiles with negative glue strengths is fundamentally limited. After t steps of assembly, $\Omega(t)$ tiles are permanently bound, unable to detach via negative glue strengths, and can only be detached by supplying external energy. The second result is a positive result that attempts to make do with this limitation: an s-space-bounded Turing machine may be simulated for arbitrarily many steps, while ensuring that no intermediate assembly grows larger than $O(s)$.

Space-bounded computation as an end goal is not the only application of negative glue strengths. We showed one example application in Section 5: assembly of a thin rectangle from a small number of tile types. There are other potential applications. Doty, Lutz, Patitz, Summers, and Woods [4] study the problem of generating uniform random distributions on the finite sets using the independent flips of a fair coin afforded by the random selection of competing tile types in the aTAM (a non-trivial problem when the cardinality of the set is not a power of the number of competing tile types), and find a tradeoff between the closeness to uniformity of the distribution obtained and the space required for sampling. They exhibit a construction imposing a perfectly uniformly distribution on the set $\{0, 1, \ldots, n-1\}$ that assembles a structure of width $\lfloor \log n \rfloor + 1$ and *expected* height at most 2, essentially implementing von Neumann's *rejection method* of flipping $\lfloor \log n \rfloor + 1$ fair coins repeatedly and stopping the first time that they encode a number smaller than n. It is very unlikely (probability at most 2^{-20}) to take more than (say) 20 attempts. But using this method in a construction such as that of [5],[12] in which many (perhaps more than 2^{20}) copies of this experiment repeat throughout assembly, could increase the likelihood of growing too high (suppose that exceeding 20 rows is too high). Even a single occurrence of a too-high subassembly will destroy the entire construction. Techniques similar to those in the present paper may be useful to augment the construction of [5] (which uses a variant of the random number selector of [4]) with negative glue strengths to implement perfectly uniform selection of random numbers, thus improving the fidelity of the simulation of [5], while providing an absolute guarantee on the space bound. More generally, negative glues could be useful in situations in which it is desirable to erase a history of failed random attempts to self-assemble a structure, so that in the final assembled structure, it appears as though a correct "guess" was made on the first attempt.

An open problem is to study the capabilities of negative glue strengths with a more constrained strength function. In our constructions, we heavily used the capability that a tile t uses one glue to repulse another tile t' and dislodge it from the assembly, and then uses that same glue to help attach a new tile. This strategy necessarily implies that the glue must interact with negative strength in the first case but positive strength in the second, forcing the glue strength function to be non-diagonal. It may be possible, however, to modify our construction to use a glue strength function that has all of its non-zero entries confined to the diagonal, so that any interaction, whether positive

[12] The main construction of [5] shows how a "universal" tile set can be constructed that can be "programmed" through appropriate selection of a seed assembly to simulate the growth of any tile assembly system in a wide class of systems termed "locally consistent" (see [5] for details). In this discussion, we are concerned only with the fact that the construction of [5] 1) requires random numbers to be generated in a bounded space at many points throughout assembly, and 2) would be improved if the distribution of these numbers were perfectly uniform instead of "close to uniform" as in [5].

or negative, is between equal glues. It seems intuitively that diagonal strength functions would be easier to implement experimentally. This remains an open problem.

It also seems intuitively more plausible to experimentally implement "non-specific" negative strengths, through some mechanism in which a protrusion on a glue pushes on a neighboring tile, but it pushes with the same strength regardless of the glue to which it is adjacent. Formally this is modeled by a strength function $str : G \times G \to \mathbb{Z}$ with the constraint that if $str(g, g') < 0$ for some $g, g' \in G$ (indicating that g pushes on g'), then for all $g'' \in G$, $str(g, g'') = str(g, g')$ (indicating that g pushes on all other glues g''). It remains an open problem to modify our main Turing machine construction to use a strength function obeying this constraint.

The previous two constraints are formally incompatible. If we assume that negative glues are easier to implement nonspecifically (due to the use of steric forces or "pushing"), and positive glues are easier to implement with a diagonal strength function (due to the use of Watson-Crick base pairing to identify matching glues), then a way to combine the previous two properties is as follows. If $str(g, g') < 0$ for some $g, g' \in G$, then for all $g'' \in G$, $str(g, g'') = str(g, g')$, and if $str(g, g') > 0$, then $g = g'$.

Another interesting constraint, considered by Patitz, Schweller, and Summers [6], is that there is only a single negative glue, and it only repels itself. This is motivated by the idea of implementing negative glues with magnets, which would push on each other but would not push or pull on the DNA sticky ends implementing positive strength glues.

Our main construction in Section 4 requires eight different nonzero strength values that range from -3 to 5. One would expect this to be more difficult to implement experimentally than a smaller range; indeed, present experimental work struggles with errors even while using only two strengths, 1 and 2. It is an open question to find the smallest interval of strengths in a temperature 2 system that achieves the result of Section 4.

Other questions related to this work include the experimental aspects of such a model, for example, how repulsive forces can be realized with DNA tiles, and how to "recycle" the junk introduced during the assembly.

Acknowledgments. The authors are grateful to Shinnosuke Seki for insightful discussions and anonymous referees for their suggestions.

References

[1] Gagan Aggarwal, Qi Cheng, Michael H. Goldwasser, Ming-Yang Kao, Pablo Moisset de Espanés, and Robert T. Schweller. Complexities for generalized models of self-assembly. *SIAM Journal on Computing*, 34:1493–1515, 2005. Preliminary version appeared in SODA 2004.

[2] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, 2001.

[3] David Doty, Lila Kari, and Benoît Masson. Negative interactions in irreversible self-assembly. In *DNA 16: Proceedings of The Sixteenth International Meeting on DNA Computing and Molecular Programming*, volume 6518 of *Lecture Notes in Computer Science*, pages 37–48. Springer, 2010.

[4] David Doty, Jack H. Lutz, Matthew J. Patitz, Scott M. Summers, and Damien Woods. Random number selection in self-assembly. In *UC 2009: Proceedings of The Eighth International*

Conference on Unconventional Computation, volume 5715 of *Lecture Notes in Computer Science*, pages 143–157. Springer, 2009.

[5] David Doty, Jack H. Lutz, Matthew J. Patitz, Scott M. Summers, and Damien Woods. Intrinsic universality in self-assembly. In *STACS 2010: Proceedings of The Twenty-Seventh International Symposium on Theoretical Aspects of Computer Science*, volume 5 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 275–286, 2010.

[6] Matthew J. Patitz, Robert T. Schweller, and Scott M. Summers. Exact shapes and turing universality at temperature 1 with a single negative glue. Technical Report 1105.1215, Computing Research Repository, 2011.

[7] John H. Reif, Sudheer Sahu, and Peng Yin. Complexity of graph self-assembly in accretive systems and self-destructible systems. In *DNA 11: Proceedings of The Eleventh International Workshop on DNA Computing*, volume 3892 of *Lecture Notes in Computer Science*, pages 257–274. Springer, 2006.

[8] Paul W. K. Rothemund and Erik Winfree. The program-size complexity of self-assembled squares (extended abstract). In *STOC 2000: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 459–468, 2000.

[9] Richard Schoen and Shing-Tung Yau. On the positive mass conjecture in general relativity. *Communications in Mathematical Physics*, 65(45):45–76, 1979.

[10] Nadrian C. Seeman. Nucleic-acid junctions and lattices. *Journal of Theoretical Biology*, 99:237–247, 1982.

[11] David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, 2008.

[12] Hao Wang. Proving theorems by pattern recognition – II. *The Bell System Technical Journal*, XL(1):1–41, 1961.

[13] Erik Winfree. *Algorithmic Self-Assembly of DNA*. PhD thesis, California Institute of Technology, June 1998.